The Annals of Statistics
2012, Vol. 40, No. 4, 1989–1996
DOI: 10.1214/12-AOS985
Main article DOI: 10.1214/11-AOS949

DISCUSSION: LATENT VARIABLE GRAPHICAL MODEL SELECTION VIA CONVEX OPTIMIZATION[1]

BY ZHAO REN AND HARRISON H. ZHOU

Yale University

1. Introduction. We would like to congratulate the authors for their refreshing contribution to this high-dimensional latent variables graphical model selection problem. The problem of covariance and concentration matrices is fundamentally important in several classical statistical methodologies and many applications. Recently, sparse concentration matrices estimation has received considerable attention, partly due to its connection to sparse structure learning for Gaussian graphical models. See, for example, Meinshausen and Bühlmann (2006) and Ravikumar et al. (2011). Cai, Liu and Zhou (2012) considered rate-optimal estimation.

The authors extended the current scope to include latent variables. They assume that the fully observed Gaussian graphical model has a naturally sparse dependence graph. However, there are only partial observations available for which the graph is usually no longer sparse. Let X be $(p+r)$-variate Gaussian with a sparse concentration matrix $S^*_{(O,H)}$. We only observe X_O, p out of the whole $p+r$ variables, and denote its covariance matrix by Σ^*_O. In this case, usually the $p \times p$ concentration matrix $(\Sigma^*_O)^{-1}$ are not sparse. Let S^* be the concentration matrix of observed variables conditioned on latent variables, which is a submatrix of $S^*_{(O,H)}$ and hence has a sparse structure, and let L^* be the summary of the marginalization over the latent variables and its rank corresponds to the number of latent variables r for which we usually assume it is small. The authors observed $(\Sigma^*_O)^{-1}$ can be decomposed as the difference of the sparse matrix S^* and the rank r matrix L^*, that is, $(\Sigma^*_O)^{-1} = S^* - L^*$. Then following traditional wisdoms, the authors naturally proposed a *regularized maximum likelihood approach* to estimate both the sparse structure S^* and the low-rank part L^*,

$$\min_{(S,L):S-L\succ 0, L\succeq 0} \operatorname{tr}\big((S-L)\Sigma^n_O\big) - \log\det(S-L) + \chi_n\big(\gamma\|S\|_1 + \operatorname{tr}(L)\big),$$

where Σ^n_O is the sample covariance matrix, $\|S\|_1 = \sum_{i,j}|s_{ij}|$, and γ and χ_n are regularization tuning parameters. Here $\operatorname{tr}(L)$ is the trace of L. The notation $A \succ 0$ means A is positive definite, and $A \succeq 0$ denotes that A is nonnegative.

There is an obvious identifiability problem if we want to estimate both the sparse and low-rank components. A matrix can be both sparse and low rank. By exploring the geometric properties of the tangent spaces for sparse and low-rank components, the authors gave a beautiful sufficient condition for identifiability, and then

Received February 2012.
[1]Supported in part by NSF Career Award DMS-06-45676 and NSF FRG Grant DMS-08-54975.